

# TMVA : merging background

## mvaBDT

day	113	120	134	160	ALL
Cut for which an event is signal-like	-0.009	-0.004	-0.006	-0.003	-0.054
Optimal cut	-0.0252	-0.0249	-0.0195	-0.0160	-0.0763

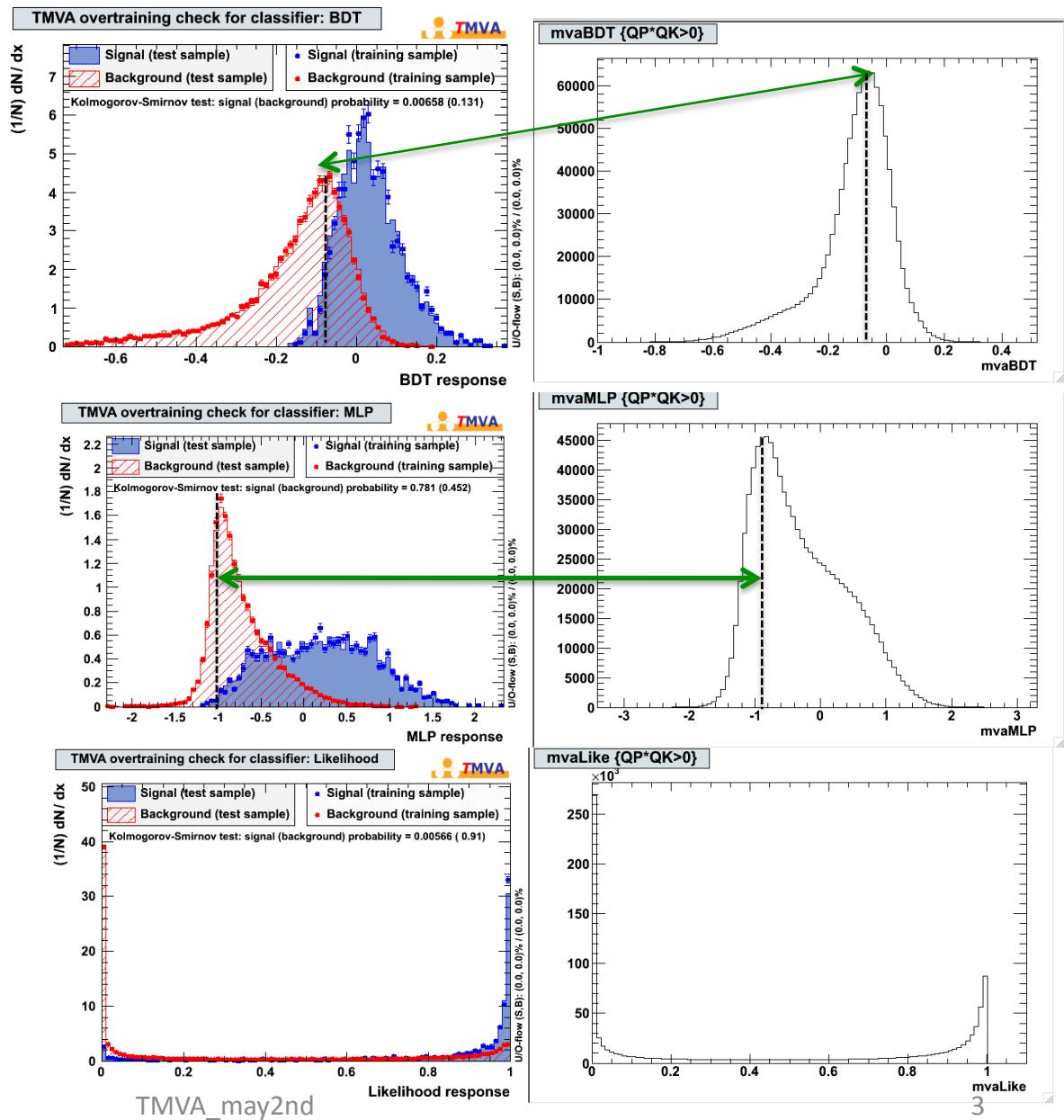
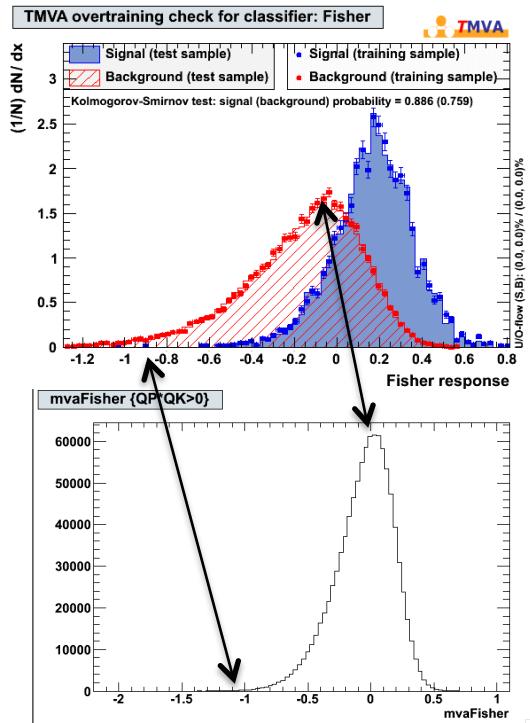
## mvaMLP

day	113	120	134	160	ALL
Cut for which an event is signal-like	-0.033	-0.052	-0.058	-0.029	-0.637
Optimal cut	-0.2420	-0.2032	-0.2124	-0.1896	-0.7566

1. The first 3 columns were the days I have looked last week :
  1. They look similar but now I see that day 160 differs
2. the results for **ALL** differs from the individual day because it is the sum of the 4 samples, meaning the size of this sample is x4 bigger
3. As a result, the value that maximizes the s/s+b is different

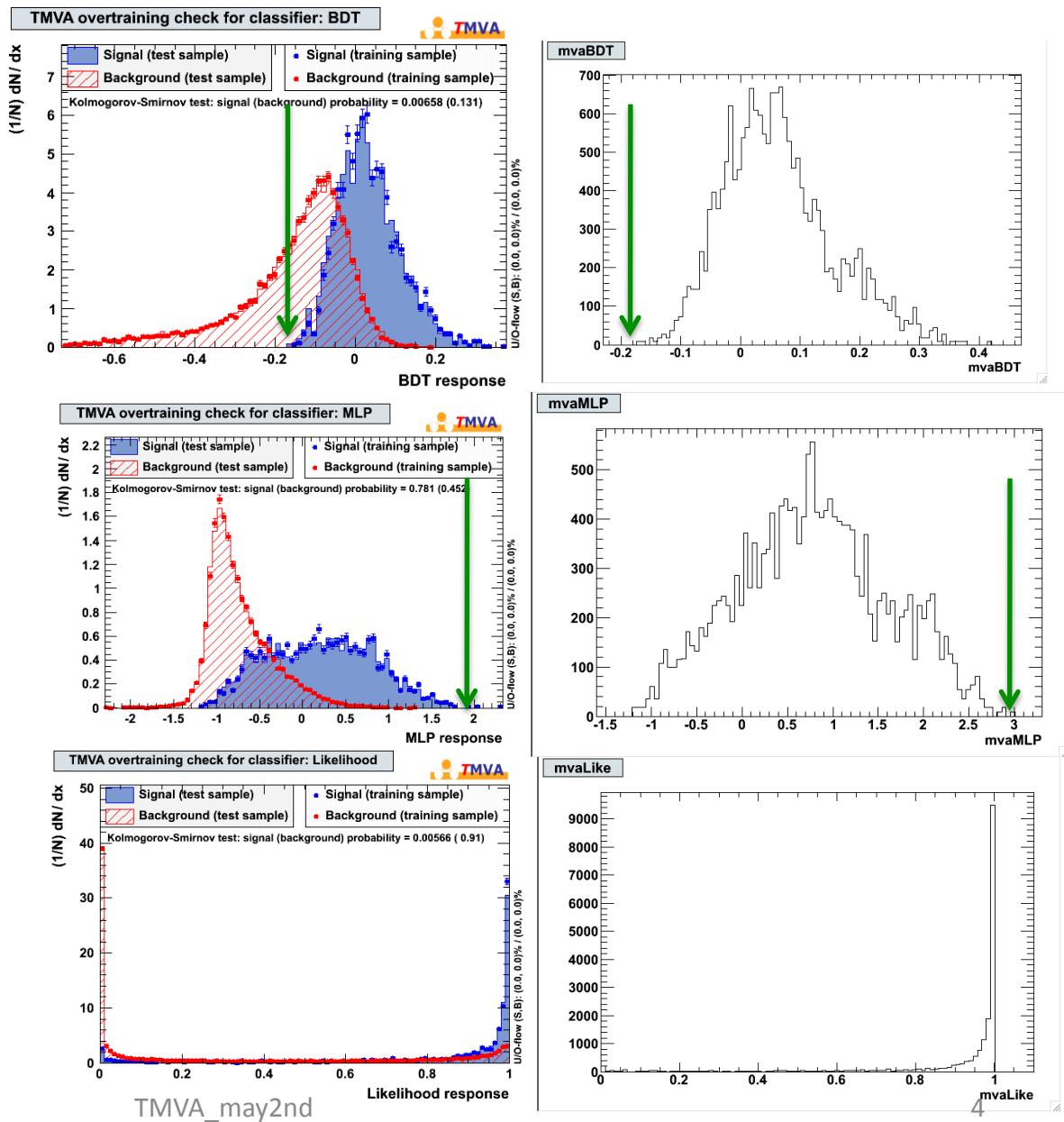
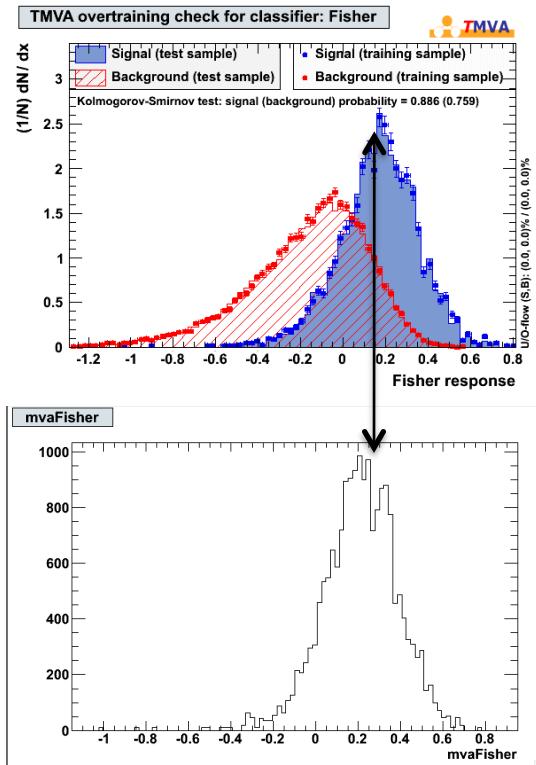
# Training day 113 ; analysis data 113 ; no pt bins

- Selecting “background” pairs
- Should match the red curve



# Training day 113 ; analysis data 113 ; no pt bins

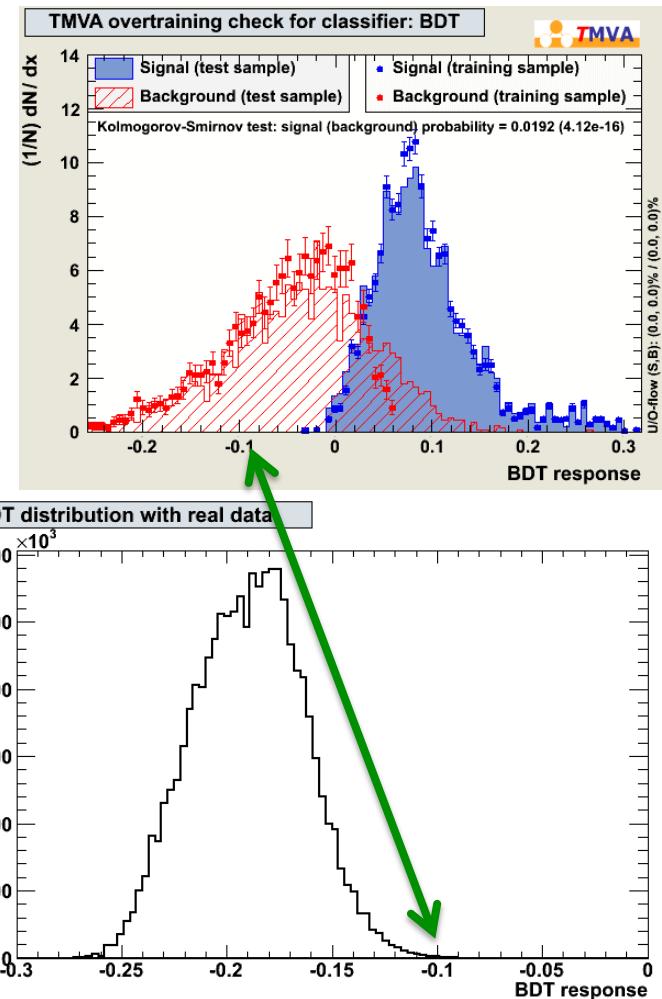
- Rerun over the signal sample after training.
- It should match the blue curve
- Fisher, BDT, Likelihood look OK
- MLP has a wider range



- The agreement is not perfect
- But compared to the plots I have shown at the analysis meeting (right), it is better
- I think the reason is because :
  - I have applied some cuts when running after the training is done (see slide 6)
  - I have used reduced variables, like  $S_L = \text{slength}/\text{dlength}$

(slide 7)

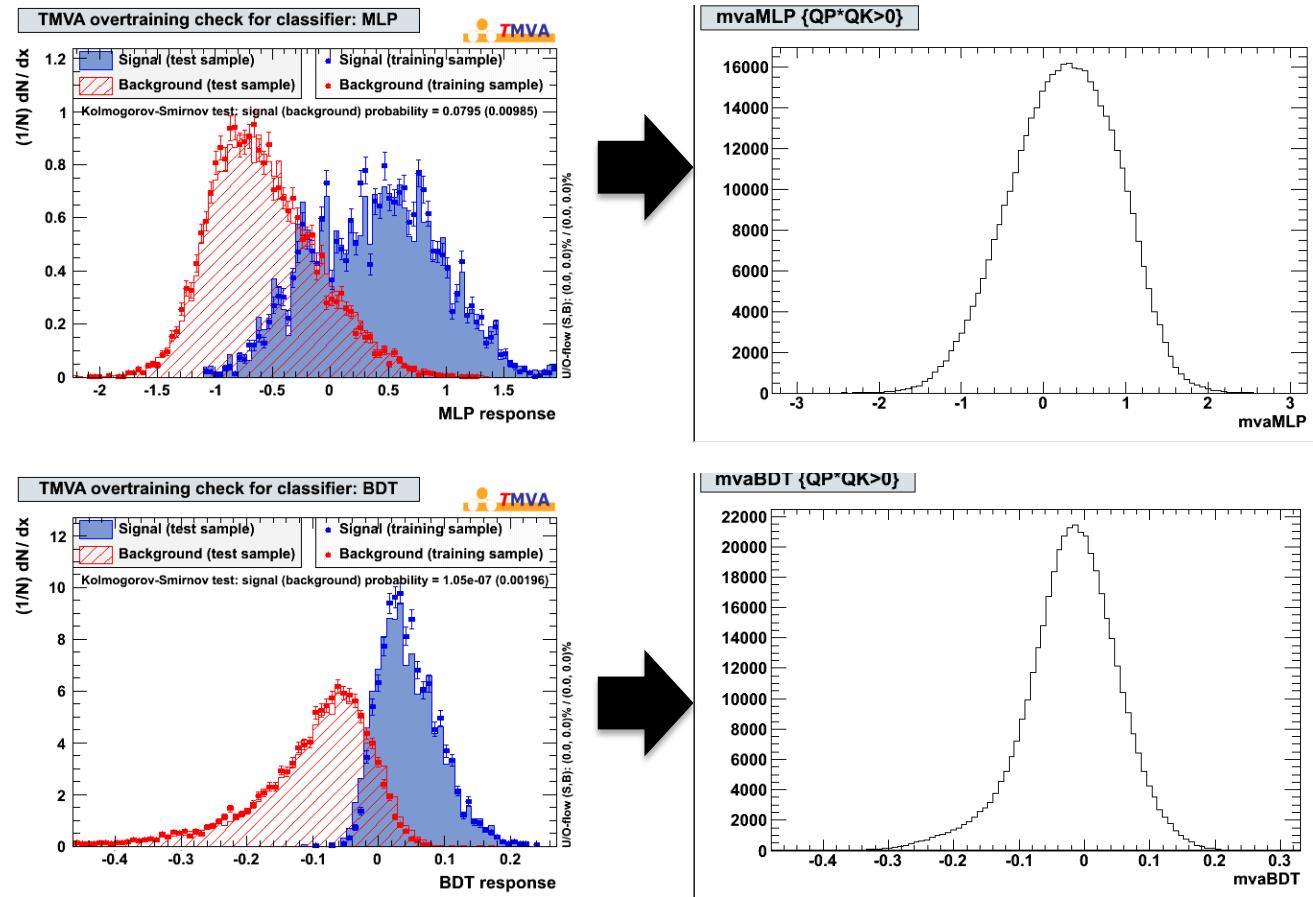
Note : reduced variable like the decay length significance might be better to use because of the Pt dependence of dca, sigmaDca



- top :
  - BDT response after training
  - A cut  $\sim 0.2$  could separate the signal from the background
- Bottom :
  - BDT response after running over real data.
  - It looks there is a shift in the shape

# Training day 113 ; analysis data 113 ; pt bins

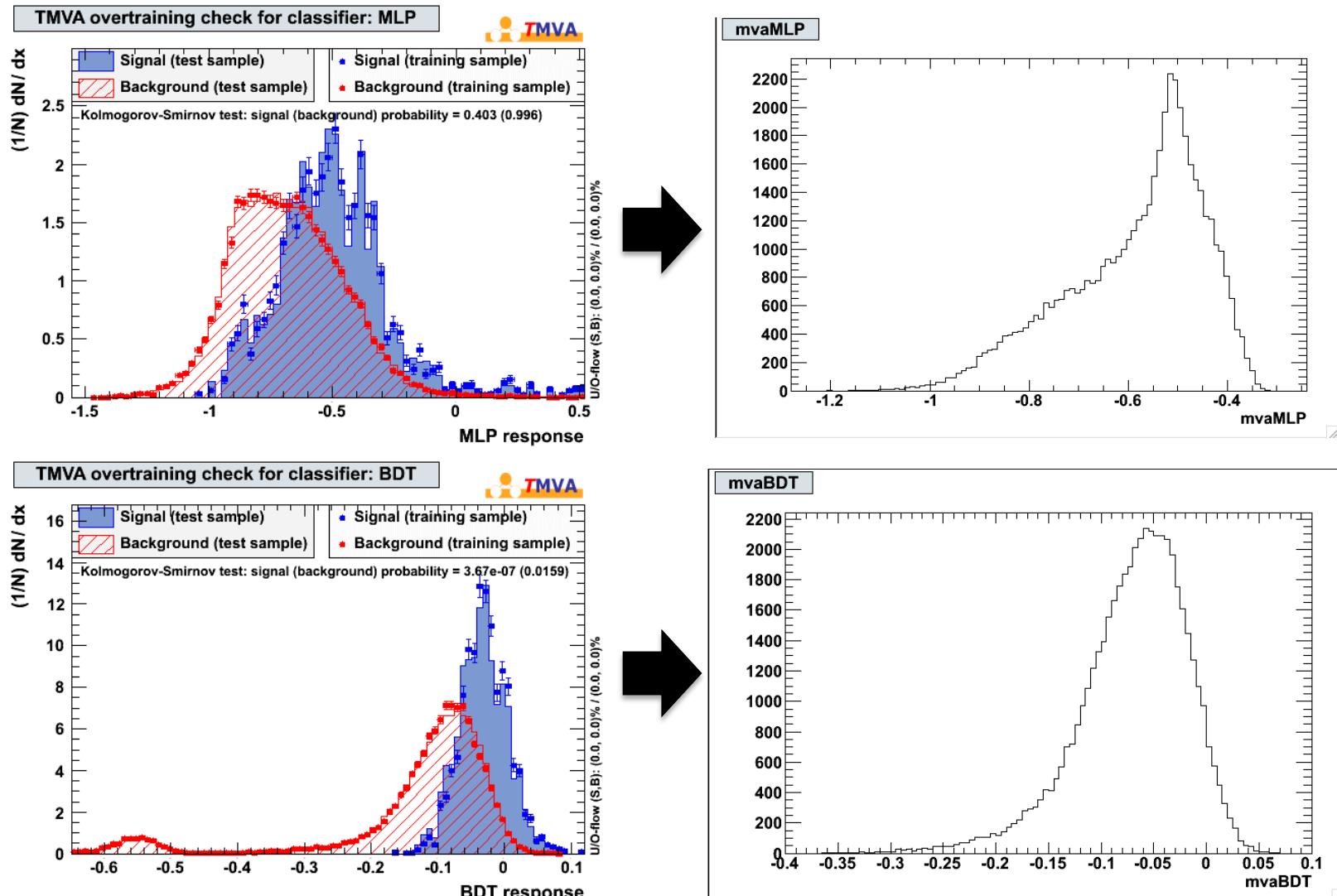
- Training is done by cutting on  $(\text{PtPion} + \text{PtKaon}) > 1.75$
- Then reading the TMVA results over the real data is also done with the same cut



$QP^*QK > 0$   
selects  
background  
pairs (like  
sign)

- Shapes and ranges of classifiers responses are different

# Training day 113 ; analysis data 113 ; reduced variables



- Shapes and ranges of classifiers responses are different

# How to make Pt bins

- p121 (TMVAguide) : The Category method allows the user to separate the training data (and accordingly the application data) into disjoint sub-populations exhibiting different properties. The separation into phase space regions is done by applying requirements on the input and/or spectator variables (variables defined as spectators are not used as direct inputs to MVA methods) .
  - Independent training in each region
  - Reduces the correlation btw variables

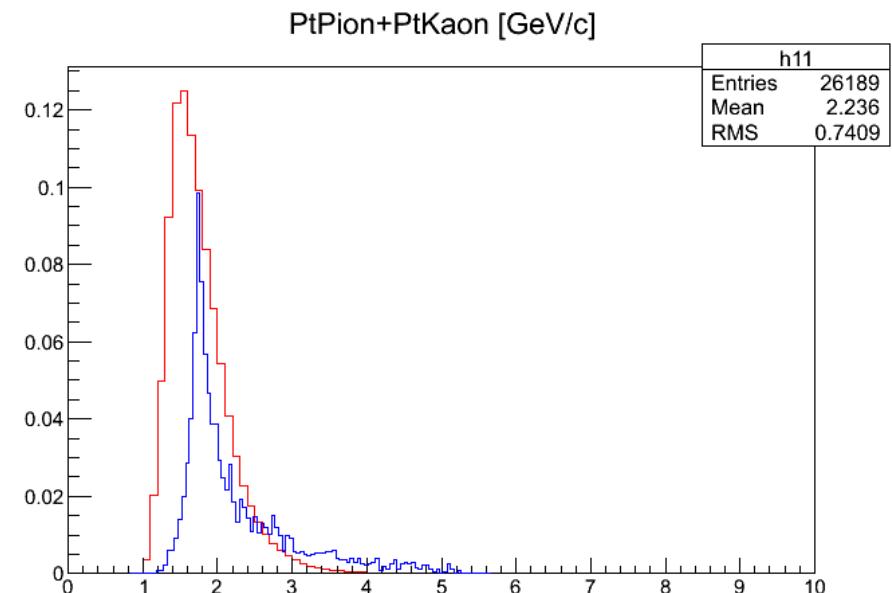
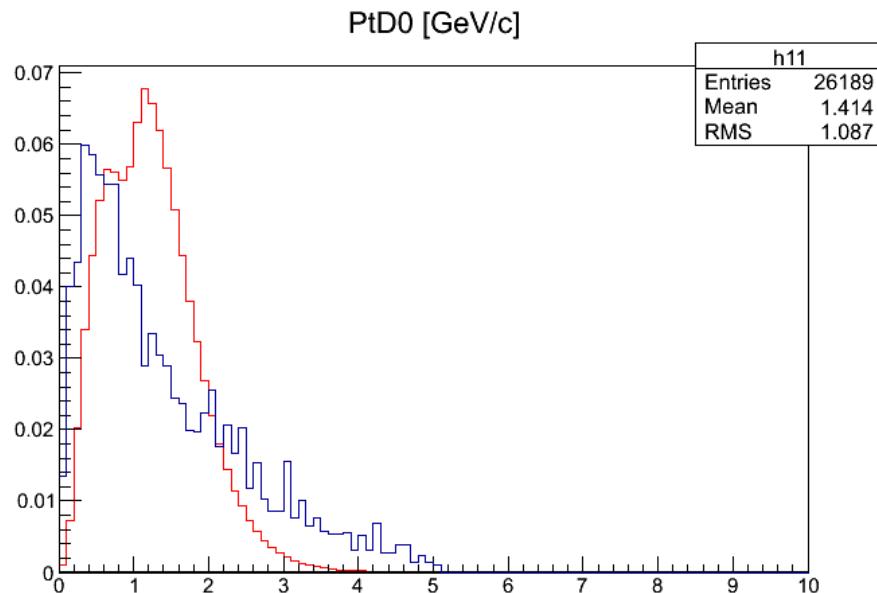
```
// Categorised classifier
TMVA::MethodCategory* mcat = 0;

// the variable sets
TString theCat1Vars = "dcaXYPion:dcaXYKaon:SigmaDcaXYPion:SigmaDcaXYKaon:dcaZPion:dcaZKaon:DcaTrackTXY:DcaTrackTZ:slength:dslength:CosPointing:thetaGJ";
TString theCat2Vars = (UseOffsetMethod ? "dcaXYPion:dcaXYKaon:SigmaDcaXYPion:SigmaDcaXYKaon:dcaZPion:dcaZKaon:DcaTrackTXY:DcaTrackTZ:slength:dslength:CosPointin
g:thetaGJ");

// the Fisher
TMVA::MethodBase* fcat = factory->BookMethod( TMVA::Types::kCategory, "FisherCat", "" );
mcat = dynamic_cast<TMVA::MethodCategory*>(fcat);
mcat->AddMethod("PtD0<=1.75",theCat1Vars, TMVA::Types::kFisher,"Category_Fisher_1","!H:!V:Fisher");
mcat->AddMethod("PtD0>1.75", theCat2Vars, TMVA::Types::kFisher,"Category_Fisher_2","!H:!V:Fisher");

// the Likelihood
TMVA::MethodBase* lcat = factory->BookMethod( TMVA::Types::kCategory, "LikelihoodCat", "" );
mcat = dynamic_cast<TMVA::MethodCategory*>(lcat);
mcat->AddMethod("PtD0<=1.75",theCat1Vars, TMVA::Types::kLikelihood,"Category_Likelihood_1","!H:!V:TransformOutput:PDFInterpol=Spline2:NSmoothSig[0]=20:NSmoothBk
g[0]=20:NSmoothBkg[1]=10:NSmooth=1:NAvEvtPerBin=50");
mcat->AddMethod("PtD0>1.75", theCat2Vars, TMVA::Types::kLikelihood,"Category_Likelihood_2","!H:!V:TransformOutput:PDFInterpol=Spline2:NSmoothSig[0]=20:NSmoothBk
g[0]=20:NSmoothBkg[1]=10:NSmooth=1:NAvEvtPerBin=50");
```

# PtD0 vs. Pt sum of daughters



- red : real data ( $K^-, \pi^+$ ) ; blue : single D0 ( $K^-, \pi^+$ )